Total Variation Distance and the Distribution of Relative Information

Sergio Verdú
Department of Electrical Engineering
Princeton University
Princeton, NJ 08544, USA
Email: verdu@princeton.edu

Abstract—We give explicit expressions, upper and lower bounds on the total variation distance between P and Q in terms of the distribution of the random variables $\log \frac{dP}{dQ}(X)$ and $\log \frac{dP}{dQ}(Y)$, where X and Y are distributed according to P and Q respectively.

I. INTRODUCTION

Two popular gauges of the distinctness between a pair of probability measures (P, Q) defined on the same measurable space $(\mathcal{A}, \mathcal{F})$ are:

Total Variation Distance

$$|P - Q| = 2 \sup_{A \in \mathcal{F}} |P(A) - Q(A)| \qquad (1)$$

Relative Entropy

$$D(P \| Q) = \mathbb{E}[\imath_{P\|Q}(X)] \qquad (2)$$

if $P \ll Q$ (which we assume throughout the paper), where we have denoted the relative information by

$$\imath_{P\|Q}(a) = \log \frac{dP}{dQ}(a) \qquad (3)$$

The most important relationship between those two quantities, sometimes referred to as "Pinsker's inequality" (see Appendix):

$$\frac{1}{2}|P - Q|^2 \log e \le D(P \| Q) \qquad (4)$$

The factor of $\frac{1}{2}$ is the best possible in the sense that

$$\inf \frac{D(P \| Q)}{|P - Q|^2} = \frac{1}{2} \log e \qquad (5)$$

In this paper, we explore upper bounds, lower bounds and exact expressions for $|P - Q|$ in terms of the distributions of the random variables

- $\imath_{P\|Q}(X)$
- $\imath_{P\|Q}(Y)$

where X and Y are random variables (from a primeval measurable space to $(\mathcal{A}, \mathcal{F})$) distributed according to P and Q, respectively. The most direct nexus between total variation distance and the distributions of the relative informations follows from the fact that the supremum in (1) is achieved by the event

$$A^* = \{a \in \mathcal{A} : \imath_{P\|Q}(a) > 0\} \qquad (6)$$

and, therefore,

$$\frac{1}{2}|P - Q| = \mathbb{P}\left[\imath_{P\|Q}(Y) \le 0\right] - \mathbb{P}\left[\imath_{P\|Q}(X) \le 0\right] \qquad (7)$$
$$= \mathbb{P}\left[\imath_{P\|Q}(X) > 0\right] - \mathbb{P}\left[\imath_{P\|Q}(Y) > 0\right] \qquad (8)$$

Furthermore, it is also easy to show

$$|P - Q| = \mathbb{E}\left[|1 - \exp(\imath_{P\|Q}(Y))|\right] \qquad (9)$$
$$= \mathbb{E}\left[|1 - \exp(-\imath_{P\|Q}(X))|\right] \qquad (10)$$

where (10) follows under the additional assumption that $Q \ll P$.

In addition to (4) and various refinements (see Appendix), some exisiting bounds in the spirit of this work are:

Theorem 1. *[5] For any $\mu > 0$,*

$$\frac{1}{2}|P - Q| \le \frac{\mu}{\log e} + \mathbb{P}[\imath_{P\|Q}(X) > \mu] \qquad (11)$$

Theorem 2. *[6] Suppose that P and Q are distributions on a finite set $\mathcal{A}$. Then*

$$\min_{a \in \mathcal{A}} Q(a) \, D(P\|Q) \le |P - Q|^2 \log e \qquad (12)$$

II. EXACT EXPRESSIONS

Theorem 3.

$$\frac{1}{2}|P - Q| = \mathbb{E}\left[[1 - \exp(-\imath_{P\|Q}(X))]^+\right] \qquad (13)$$
$$= \mathbb{E}\left[[1 - \exp(\imath_{P\|Q}(Y))]^-\right]. \qquad (14)$$
$$= \mathbb{E}\left[[1 - \exp(\imath_{P\|Q}(Y))]^+\right] \qquad (15)$$
$$= \int_0^1 \mathbb{P}\left[\imath_{P\|Q}(Y) < \log \beta\right] d\beta \qquad (16)$$
$$= \int_0^1 \mathbb{P}\left[\imath_{P\|Q}(X) > \log \frac{1}{\beta}\right] d\beta, \qquad (17)$$
$$= \mathbb{E}\left[[1 - \exp(-\imath_{P\|Q}(X))]^-\right] \qquad (18)$$

where $[a]^+ = a\,1\{a > 0\}$, $[a]^- = a\,1\{a < 0\}$, and (18) follows under the additional assumption that $Q \ll P$.

Proof: By change of measure, we obtain (13):

$$\mathbb{E}\left[[1 - \exp(-\imath_{P\|Q}(X))]^+\right]$$
$$= \mathbb{E}\left[(1 - \exp(-\imath_{P\|Q}(X))1\{\imath_{P\|Q}(X) > 0\}\right] \qquad (19)$$
$$= \mathbb{P}\left[\imath_{P\|Q}(X) > 0\right] - \mathbb{P}\left[\imath_{P\|Q}(Y) > 0\right] \qquad (20)$$

Similarly, we obtain (14), and since $[a]^+ + [a]^- = |a|$, (15) follows from (9).

We abbreviate $Z = \exp(\imath_{P\|Q}(Y))$. To show (16) note that

$$\mathbb{E}\left[(Z-1)\mathbf{1}\{Z > 1\}\right] = \int_1^\infty \mathbb{P}\left[Z > \beta\right]\mathrm{d}\beta \qquad (21)$$

$$= \int_0^1 \mathbb{P}\left[Z < \beta\right]\mathrm{d}\beta \qquad (22)$$

$$= \mathbb{E}\left[(1-Z)\mathbf{1}\{Z < 1\}\right] \qquad (23)$$

where

- (22) $\Longleftarrow$ adding $\int_0^1 \mathbb{P}[Z > \beta]\,\mathrm{d}\beta$ to both sides makes them equal to 1 since $\mathbb{E}[Z] = 1$;
- (21) and (23) $\Longleftarrow$ if $V \geq 0$, then $\mathbb{E}[V] = \int_0^\infty \mathbb{P}[V > t]\mathrm{d}t$;

Finally, to show (17), consider

$$\int_0^1 \mathbb{P}\left[\imath_{P\|Q}(X) > \log\frac{1}{\beta}\right]\mathrm{d}\beta$$

$$= \int_0^1 \mathbb{E}\left[\mathbf{1}\left\{Z > \frac{1}{\beta}\right\}Z\right]\mathrm{d}\beta \qquad (24)$$

$$= \int_0^1 \int_0^\infty \mathbb{P}\left[\mathbf{1}\left\{Z > \frac{1}{\beta}\right\}Z > t\right]\mathrm{d}t\,\mathrm{d}\beta \qquad (25)$$

$$= \int_0^1 \frac{1}{\beta}\mathbb{P}\left[Z > \frac{1}{\beta}\right]\mathrm{d}\beta + \int_0^1 \int_{1/\beta}^\infty \mathbb{P}[Z > t]\,\mathrm{d}t\,\mathrm{d}\beta \qquad (26)$$

$$= \int_0^1 \frac{1}{\beta}\mathbb{P}\left[Z > \frac{1}{\beta}\right]\mathrm{d}\beta + \int_1^\infty \left(1 - \frac{1}{\beta}\right)\mathbb{P}[Z > \beta]\,\mathrm{d}\beta \qquad (27)$$

$$= \int_1^\infty \mathbb{P}[Z > \beta]\,\mathrm{d}\beta \qquad (28)$$

$$= \int_0^1 \mathbb{P}[Z \leq \beta]\,\mathrm{d}\beta \qquad (29)$$

where

- (24) $\Longleftarrow$ change of measure;
- (27) $\Longleftarrow$ swapping the order of integration in the second integral;
- (28) $\Longleftarrow$ change of variable of integration $\frac{1}{\beta} \leftarrow \beta$ in the first integral;
- (29) $\Longleftarrow$ $\mathbb{E}[Z] = 1$.

Finally, under the additional assumption $Q \ll P$, (18) follows from (10) and (13). ∎

III. UPPER BOUNDS ON $|P - Q|$

Often, the exact expressions for $|P - Q|$ in (7)-(9) are hard to evaluate precisely. In particular, except in elementary settings, it is not always easy to get a handle on *both* relative information distributions as required in (7). Extremely useful as it is, (4) is inoperative when the relative entropy is larger than 2 nats. Moreover, sometimes to show convergence in total variation distance, instead of attempting to show the stronger result in relative entropy, it is simpler to use a bound such as (11), for an appropriately shrinking sequence of μ.

Our first new upper bound refines the bound by Pinsker (49), and in fact it can be sharper than (4).

Theorem 4.

$$|P - Q|\log e \leq D(P\|Q) + \mathbb{E}\left[|\imath_{P\|Q}(X)|\right] \qquad (30)$$

Proof: For all $z \in [-\infty, +\infty]$,

$$(1 - \exp(-z))\mathbf{1}\{z > 0\} \leq \frac{z}{\log e}\mathbf{1}\{z > 0\} \qquad (31)$$

Letting $z \leftarrow \imath_{P\|Q}(X)$, and taking expectation of both sides with respect to X, the right side of (31) becomes one half of the right side of (30) (in nats), while the expectation of the left side of (31) is equal to $\frac{1}{2}|P - Q|$ as we showed in (13). ∎

The following result strengthens Theorem 1:

Theorem 5.

$$\frac{1}{2}|P - Q| \leq \min_{\beta_1 < \beta_0 \leq 1}\left\{(1 - \beta_0)\,\mathbb{P}\left[\imath_{P\|Q}(X) \geq 0\right] + (\beta_0 - \beta_1)\,\mathbb{P}\left[\imath_{P\|Q}(X) > \log\frac{1}{\beta_0}\right]\right\} \qquad (32)$$

and

$$\beta_1^{-1} = \sup_{a \in \mathcal{A}}\frac{\mathrm{d}P}{\mathrm{d}Q}(a) \qquad (33)$$

with $\beta_1 = 0$ *if the relative information is unbounded from above.*

Furthermore,

$$\frac{1}{2}|P - Q| \leq \min_{\beta_2 < \beta_0 \leq 1}\left\{(1 - \beta_0)\,\mathbb{P}\left[\imath_{P\|Q}(Y) \leq 0\right] + (\beta_0 - \beta_2)\,\mathbb{P}\left[\imath_{P\|Q}(Y) < \log\beta_0\right]\right\} \qquad (34)$$

and

$$\beta_2 = \inf_{a \in \mathcal{A}}\frac{\mathrm{d}P}{\mathrm{d}Q}(a) \qquad (35)$$

Proof: Since the integrand in (17) is monotonically increasing with β, we may upper bound it by $\mathbb{P}\left[\imath_{P\|Q}(X) > \log\frac{1}{\beta_0}\right]$ when $\beta_1 \leq \beta \leq \beta_0$ and by $\mathbb{P}\left[\imath_{P\|Q}(X) \geq 0\right]$ when $\beta_0 < \beta \leq 1$. The same reasoning applied to (16) yields (34). ∎

Note that we can get a simple upper bound on $|P - Q|$ by simply dropping the second term in the right side of (8). A tighter result ensues by choosing the left-most value in (32):

$$\frac{1}{2}|P - Q| \leq (1 - \beta_1)\,\mathbb{P}\left[\imath_{P\|Q}(X) \geq 0\right]. \qquad (36)$$

IV. LOWER BOUNDS ON $|P - Q|$

For any $0 < \tau$, we can find a pair of distributions such that $|P - Q| < \tau$ and $D(P\|Q) = +\infty$. Consequently, we cannot find a general lower bound on the total variation distance as a function of the relative entropy only. Nevertheless, it is indeed possible to make some progress in this direction. The first lower bound we offer is the counterpart of Theorem 4.

Theorem 6. *If* $P \lll\ggg Q$, *then*

$$|P - Q|\log e \geq \mathbb{E}\left[|\imath_{P\|Q}(X)|\right] - D(P\|Q) \qquad (37)$$

Proof: We reason in parallel to the proof of Theorem 4. For all $z \in [-\infty, +\infty]$,

$$[1 - \exp(-z)]^- \log e \geq [z]^- \qquad (38)$$

Letting $z = \imath_{P\|Q}(X)$, and taking expectation of both sides with respect to X, the right side of (38) becomes one half of the right side of (37), while the expectation of the left side of (38) is equal to $\frac{1}{2}|P - Q| \log e$ because of (18). ∎

If the pair of distributions is such that their relative information is bounded, small total variation distance does imply small relative entropy, as we see next.

Theorem 7. *With $\beta_1 \leq 1$ defined in* (33), *we have*

$$\frac{1}{2}|P - Q| \geq \frac{1 - \beta_1}{\log \frac{1}{\beta_1}} D(P \| Q) \qquad (39)$$

$$\geq \frac{\sqrt{\beta_1}}{\log e} D(P \| Q) \qquad (40)$$

Proof: The function $\frac{z \log z}{z-1}$ is monotonically increasing for $z > 1$. Therefore, for $0 \leq z \leq \beta_1^{-1}$,

$$\frac{1 - \beta_1}{\log \frac{1}{\beta_1}} z \log z \leq [z - 1]^+ \qquad (41)$$

Substituting $z \leftarrow \exp(\imath_{P\|Q}(Y))$ and taking expectations of both sides of (41), we obtain (39) because of (15). Furthermore, (40) follows because for $0 < x < 1$,

$$\sqrt{x} \leq \frac{x - 1}{\log_e x} \qquad (42)$$

∎

Another lower bound on total variation distance based on the distribution of the relative information is given by the following result.

Theorem 8.

$$|P - Q| \geq \sup_{\eta > 0} (1 - \exp(-\eta)) \, \mathbb{P}\left[|\imath_{P\|Q}(X)| > \eta\right] \qquad (43)$$

Proof: For any $\eta > 0$,

$$|P - Q| = \mathbb{E}\left[|\exp(-\imath_{P\|Q}(X)) - 1|\right] \qquad (44)$$

$$\geq \mathbb{E}\left[|\exp(-\imath_{P\|Q}(X)) - 1| \, 1\{|\imath_{P\|Q}(X)| > \eta\}\right] \qquad (45)$$

$$\geq (1 - \exp(-\eta)) \, \mathbb{P}\left[|\imath_{P\|Q}(X)| > \eta\right] \qquad (46)$$

because if $|z| > \eta$, then $|\exp(-z) - 1| > 1 - \exp(-\eta)$. ∎

If either $\mathbb{P}\left[\imath_{P\|Q}(X) > \log \frac{1}{\beta}\right]$ or $\mathbb{P}\left[\imath_{P\|Q}(Y) < \log \beta\right]$ are known for at least one value of β, then we get the following lower bounds on $|P - Q|$ simply by invoking the exact expressions for as a function of the distribution of the relative information in Theorem 3.

Theorem 9. *For any $0 \leq \beta_0 \leq 1$,*

$$|P - Q| \geq 2(1 - \beta_0)\mathbb{P}\left[\imath_{P\|Q}(X) > \log \frac{1}{\beta_0}\right] \qquad (47)$$

$$|P - Q| \geq 2(1 - \beta_0)\mathbb{P}\left[\imath_{P\|Q}(Y) < \log \beta_0\right] \qquad (48)$$

V. APPENDIX: "PINSKER'S INEQUALITY"

The folklore in information theory is that the bound (4) is due to Pinsker [1], albeit with a suboptimal constant. Leaving aside the inessential aspect that Pinsker considered information density in lieu of relative information (i.e. $P_{XY} \ll P_X \times P_Y$ instead of $P \ll Q$), he proved (in nats)

$$\frac{1}{2}|P - Q| \leq \mathbb{E}\left[|\imath_{P\|Q}(X)|\right] \qquad (49)$$

$$\mathbb{E}\left[|\imath_{P\|Q}(X)|\right] \leq D(P\|Q) + 10\sqrt{D(P\|Q)} \qquad (50)$$

Capitalizing on the fact that $|P - Q| \leq 2$, it can be shown that putting together (49) and (50) yields

$$\frac{D(P\|Q)}{|P - Q|^2} > \frac{\log e}{408} \qquad (51)$$

The proof of (4) is due to Csiszár [2] and Kullback [3], with Kemperman [4] independently a bit later. Further refinements of (4) can be found in [7] and [8] which gives the following parametric expression (in nats) for the minimal relative entropy compatible with a given total variation distance (which was originally considered in [9]):

$$L(x(t)) = \log\left(\frac{t}{\sinh(t)}\right) + t \coth(t) - \frac{t^2}{\sinh^2(t)} \qquad (52)$$

$$x(t) = t - \frac{1}{t}\left(t \coth(t) - 1\right)^2 \qquad (53)$$

ACKNOWLEDGMENT

REFERENCES

[1] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco: Holden-Day, 1964, originally published in Russian in 1960.

[2] I. Csiszár, "Information-type measures of difference of probability distributions and indirect observations," *Studia Scientiarum Mathematicarum Hungarica*, vol. 2, pp. 299–318, 1967.

[3] S. Kullback, "A lower bound for discrimination information in terms of variation," *IEEE Trans. Information Theory*, vol. 13, pp. 126–127, Jan. 1967.

[4] J. H. B. Kemperman, "On the optimal rate of transmitting information," *Annals Mathematical Statistics*, vol. 40, pp. 2156–2177, Dec. 1969.

[5] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Information Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[6] I. Csiszár and Z. Talata, "Context tree estimation for not necessarily finite memory processes, via BIC and MDL," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1007–1016, 2006.

[7] E. Ordentlich and M. Weinberger, "A distribution dependent refinement of Pinsker's inequality," *IEEE Trans. Information Theory*, vol. 51, no. 5, pp. 1836–1840, May 2005.

[8] A. A. Fedotov, P. Harremoes, and F. Topsoe, "Refinements of Pinsker's inequality," *IEEE Transactions on Information Theory*, vol. 49, no. 6, pp. 1491–1498, 2003.

[9] I. Vajda, "Note on discrimination information and variation," *IEEE Trans. Information Theory*, vol. 16, pp. 771–773, Nov. 1970.